Published monthly by the
VIRGINIA DEPARTMENT OF HEALTH
Office of Epidemiology
P.O. Box 2448
Richmond, Virginia 23218
http://www.vdh.virginia.gov
Telephone: (804) 864-8141

PRESORTED
STANDARD
U.S. POSTAGE
PAID
Richmond, Va.
Permit No. 591

VIRGINIA
EPIDEMIOLOGY BULLETIN

Robert B. Stroube, M.D., M.P.H., Health Commissioner
Carl W. Armstrong, M.D., State Epidemiologist
Christopher Novak, M.D., M.P.H., Editor
Vickie L. O'Dell, Layout Editor

Volume 107, No.2

Disease Cluster Investigations

Introduction

The Centers for Disease Control and Prevention (CDC) defines a cluster as "an unusual aggregation, real or perceived, of health events that are grouped together in time and space."[1] Although the term can be applied on a very large scale, it is more commonly used to describe elevated rates in localized areas, such as a town or neighborhood.[2] Individuals may suspect a cluster when they believe that an unusual amount of illness has occurred in family members, friends, neighbors, or coworkers.[1] Patients often initially express these concerns about illness in their community to a healthcare professional. As such, healthcare professionals should be aware of some of the issues related to investigating clusters of illness. This article briefly reviews the processes, and the challenges, involved in evaluating clusters.

Disease Clusters

Disease clusters occur; some are due to chance, others are due to common exposures or genetic predispositions. Obvious clusters include outbreaks of gastroenteritis (e.g., salmonellosis, norovirus) or respiratory illness (e.g., influenza, tuberculosis). Studies of these clusters can develop important information on disease prevention.

Classic cancer clusters that have provided new scientific information include:

- Scrotal cancer among London chimney sweeps in the 18th century due to exposure to coal soot;
- Osteosarcoma of the jaw in women employed as watch dial painters following exposure to radium-containing paint;
- Pleural mesothelioma among asbestos workers in London; and,
- Angiosarcoma of the liver among chemical workers exposed to vinyl chloride monomer.

Another cluster investigation, of Kaposi's sarcoma and *Pneumocystis jirovecii* pneumonia in gay men, led to the discovery of the human immunodeficiency virus (HIV).[3]

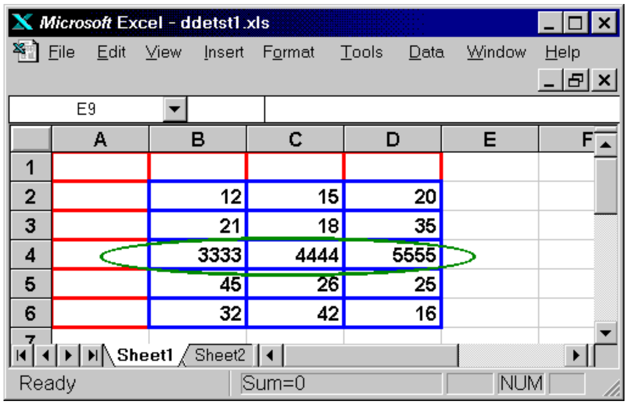
Most disease cluster studies that have yielded etiologic information have been studies of occupational, drug-induced, or infectious pathogenic exposure.[4] Studies of the typically low intensity exposures caused by environmental factors (e.g., high-tension power lines, nuclear facilities, hazardous waste dumps) have been less productive. These have not generally provided new information about the causes or prevention of illness, nor have they convincingly identified a reason for apparent clustering.[3]

Cluster Investigations

Initial Response

The first step for health department staff who are contacted regarding a suspected cluster of illness is to collect additional information. A line listing is developed for each known case in the cluster, including illness (e.g., type of cancer), date of diagnosis, age at diagnosis, and any known risk factors.[1] Other important information are the geographic area and time period of concern, the suspected exposure(s), duration of exposure, how the caller learned about the alleged cluster, and why the caller is concerned. Contact information on the person reporting the cluster is also important—while callers sometimes request anonymity, this may limit the ability to follow-up and gather additional information.[1]

Further discussion with the reporting individual explores the individual's initial impressions and understanding of disease clustering. Very often, the discussion resolves the concerns. However, if a problem may exist (e.g., the cluster contains similar illnesses, the exposure is biologically plausible, the number of cases is large enough to evaluate, and the geographic area can be defined) then additional investigation may be warranted.[1]



Basic Investigation

Even when cases occur randomly in a population, some amount of clustering may occur simply by chance.[3,5] The fact that the boundaries (or time period) of a suspected cluster are defined based on when and where the cases actually occurred increases the likelihood that random variation will be described as a cluster.[2,3] This is comparable to a "sharpshooter" who first fires randomly at a wall and then draws a bull's-eye around a cluster of bullet holes.[3]

Further investigation of a cluster requires consideration of the geographic area involved, the appropriate time period for study, the potentially exposed population, and the type(s) of illness to be included in the analysis.[1] This leads to developing a case definition (i.e., the criteria that guide inclusion as a case).[5] An appropriate reference population, such as a neighboring health district, then needs to be selected for comparison[1]—this is often a significant challenge and study limitation. Sources of data for the comparison could include reportable disease data, disease registries, death certificates, national surveys, medical records, etc. Comparisons of incidence rates and mortality ratios (adjusted for particular factors, such as age, gender, or race/ethnicity, as necessary) help to determine if the observed situation is significantly different from what would be expected.[1,5] Epidemiologists can then use statistical tests to quantify how likely an observed difference is due to chance.[5]

A suspected cluster is more likely to be a true cluster if it involves:

- A large number of cases of one type of illness, rather than several different types;
- Biological plausibility and adequate latency for the reported illness;
- A rare type of illness;
- A common illness in an atypical demographic group;
- Specific exposure to a known agent; and/or,
- An elevated ratio of observed/expected confirmed cases.[1,4]

If this preliminary evaluation suggests that an excess of illness is present and that a condition of biologic and public health importance could exist, then the need for further assessment will be determined.[1] This requires carefully balancing the scientific evidence with the level of community concern. Sometimes, public pressure alone can lead public health officials to undertake an investigation that they do not believe is warranted. However, additional studies after experts have concluded that nothing out of the ordinary has occurred are unlikely to produce useful results.[2]

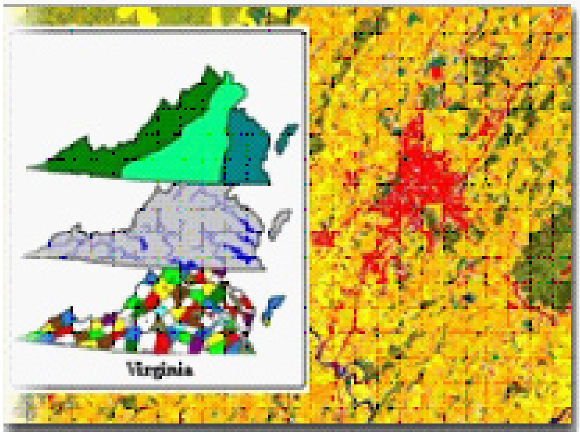
Feasibility Review

To determine if a more in-depth study may be necessary, additional detailed information on each case may be needed. Often, the health department works with the concerned individual reporting the cluster to further identify cases and to collect the information. Important information to gather includes data on the illness, exposure histories (e.g., smoking, occupation, residence), and family history.[1]

Based on the available data, a review of the literature, and consultation with other experts (e.g., Centers for Disease Control and Prevention), the feasibility of a study to identify an etiologic agent could be considered.[1] In general, these kinds of studies may be technically difficult, time consuming, and very resource intensive, so they are only done when the potential benefits are clear.

Epidemiologic Study

The purpose of an epidemiologic study is to identify a potential disease-exposure relationship. The approach taken depends on the nature of the cluster and the data that may be available. A case-control or cohort study could be used to evaluate the association between specific risk factors and the observed illness.[1] Cluster analysis techniques have also been advanced by the use of Geographic Information System (GIS) tools that can depict and display potential clusters in visually compelling ways. However, even with these tools, if the initial investigations did not indicate a likely cause then extensive follow-up investigations can take years to complete but generally have inconclusive results.[4]



Cluster Challenges

The public perception of clusters, as popularized by movies such as *Erin Brockovich*, is that any collection of people diagnosed with an illness may represent a mini-epidemic. This image dramatically underestimates the complexities of disease cluster investigations and does not address the fact that most perceived clusters include different types of illnesses, cases with little connection to the community, and/or cases that occurred over a longer time period than appreciated.[3]

While epidemiologists seek to understand the causes of disease, and to develop ways to prevent illness, there are also limitations.[1] Investigation of some clusters, such as acute infectious diseases that are relatively limited in time and space, can be very productive. Other clusters may be complicated by diseases with long latency periods, multiple causes and interactive effects, bias (e.g., recall bias), the effects of individual susceptibility and exposure history, migration, limited case numbers and/or baseline information, insufficient environmental data, and/or poor data quality.[4,5]

Sometimes a suspected cluster has enough cases for study, but a greater than expected number of cases cannot be demonstrated. And even when an investigation documents that a given clustering is "statistically significant", this does not necessarily mean that there is any single external cause or hazard that can be identified. A 'confirmed' cluster could be the result of chance, miscalculation of the expected number of cases, or differences in the case definition between observed cases and expected cases. Sometimes epidemiologists do find a true excess of cases, but they cannot find an explanation for it.[5] Or, a study may identify factors associated with illness in a group of people—but these results generally do not enable determining the specific cause of disease in an individual.[3]

In other situations, community members are concerned about a specific exposure (e.g., an industrial or agricultural pollutant). In order for an agent to be linked to a cluster, it must not only be present in the environment, but it must have come into contact with persons through a means that would present a risk and in a dose sufficient to cause illness. The levels of exposure to environmental agents in non-occupational settings are much lower and more difficult to assess than for workplaces. Even when exposure levels exceed environmental standards, the expected increase in risk from community exposures would generally be detectable only in very large populations. Epidemiologic methods that can provide evidence of association in large studies may have limited utility in evaluating localized clustering.[3]

Finally, no matter how many environmental causes are ruled out, it will always remain possible that some unknown agent is the cause of a cluster. Thus, investigations of environmental causes for a cluster can be extended almost indefinitely as more and more agents are examined. As a result, public health officials often find themselves at a point where they cannot, no matter how much they investigate, prove that an elevated rate *is* a coincidence, even if that is the real explanation.[2]

Conclusions

Identifying the cause of a disease cluster is usually made through knowledge of disease patterns and statistics. Clusters of acute illness (e.g., infectious disease, injuries) are generally easier to investigate due to the short time periods involved, availability of testing, etc. In contrast, the vast majority of apparent chronic disease clusters are chance events and are not due to an identifiable common cause.[1]

The decision as to whether or not to conduct further investigation of a cluster is often difficult. There are many more reports of suspected clusters of illness than can or should be investigated extensively.[3] Public health should respond to community concerns, document the facts of what has happened (and thereby minimize the influence of rumor), and assist the community in determining and implementing the appropriate response (e.g., environmental monitoring for an identifiable source of contamination).[3,4] However, environmental measurements rarely resolve controversy about the cause of a cluster and will not, by themselves, provide scientifically convincing evidence linking a cluster to an environmental exposure.[3]

To some it may appear negligent not to explore every possible explanation for an apparent cluster. However, the desire to "leave no stone unturned" is not in itself a sufficient reason to conduct extensive environmental monitoring or medical testing. Professional judgment about the likelihood of whether further investigation will be informative should help to guide health officials and communities confronting these difficult situations.[3]

Known or suspected outbreaks of any condition must be reported to the local health department. However, healthcare profession-

Cases of Selected Notifiable Diseases Reported in Virginia*

Disease	Total Cases Reported, January 2007: State	Regions: NW	N	SW	C	E	Total Cases Reported Statewide, January: This Year	Last Year	5 Yr Avg
AIDS	27	7	14	2	0	4	27	24	40
Campylobacteriosis	16	1	7	2	5	1	16	20	11
Chickenpox	1	0	1	0	0	0	1	5	17
E. coli, Shiga toxin-producing	6	3	2	0	1	0	6	0	1
Giardiasis	23	7	10	0	5	1	23	10	11
Gonorrhea	311	18	11	51	103	128	311	376	674
Group A Strep, Invasive	5	0	0	2	3	0	5	8	3
Hepatitis, Viral									
A	3	1	1	1	0	0	3	0	2
B, acute	2	0	0	0	0	2	2	0	3
C, acute	0	0	0	0	0	0	0	0	0
HIV Infection	29	2	6	3	2	16	29	45	48
Lead in Children†	12	3	0	4	1	4	12	20	21
Legionellosis	2	1	1	0	0	0	2	1	1
Lyme Disease	7	0	6	0	1	0	7	0	<1
Measles	0	0	0	0	0	0	0	0	0
Meningococcal Infection	2	1	0	1	0	0	2	2	1
Pertussis	9	3	3	1	0	2	9	0	1
Rabies in Animals	30	9	4	7	8	2	30	43	29
Rocky Mountain Spotted Fever	1	1	0	0	0	0	1	0	0
Rubella	0	0	0	0	0	0	0	0	0
Salmonellosis	44	9	11	7	9	8	44	11	19
Shigellosis	2	0	1	1	0	0	2	0	9
Syphilis, Early§	38	2	17	2	12	5	38	23	11
Tuberculosis	0	0	0	0	0	0	0	2	4

Localities Reporting Animal Rabies This Month: Arlington 1 raccoon; Augusta 1 cow, 1 fox; Bedford 2 raccoons; Buchanan 1 raccoon; Carroll 1 skunk; Charles City 1 dog; Fairfax 1 fox, 1 raccoon, 1 skunk; Fauquier 1 skunk; Floyd 1 raccoon; Fluvanna 1 fox; Hanover 3 raccoons, 3 skunks; Isle of Wight 1 skunk; Mecklenburg 1 skunk; Radford 1 skunk; Rappahannock 1 raccoon; Rockbridge 2 skunks; Spotsylvania 1 fox, 1 skunk; Virginia Beach 1 raccoon; Wythe 1 skunk.
Toxic Substance-related Illnesses: Adult Lead Exposure 5; Mercury Exposure 2; Pneumoconiosis 3.

*Data for 2007 are provisional. †Elevated blood lead levels ≥10µg/dL. §Includes primary, secondary, and early latent.

als may assess a patient's concerns, and assist them in understanding disease clusters, by:

- Encouraging patients to be skeptical. Community members who raise concerns about possible clusters will frequently explain themselves in terms of a "common sense" feeling that something is wrong.[2] However, "common sense" does not trump science in this area.
- Encouraging patients to know their sources: activists and experts often disagree when it comes to some clusters.
- Encouraging patients to remember the bull's-eye effect: check for bias in the way that statistics are organized or calculated.
- Clearly identifying helpful actions that people can take to reduce their risk of illness.

These steps help to maximize the efficacy of resources by allowing public health officials and scientists to direct and manage the details of research efforts. However, individuals who continue to have concerns about illness in their community may be referred to their local health department for assistance.

References

1. Virginia Department of Health, Virginia Cancer Registry. Protocol for the Investigation of Reported Cancer Clusters, March 7, 2002.
2. Robinson D. Accessed 2/2/07 at www.medscape.com/viewarticle/442554_1.
3. Thun MJ, Sinks T. *CA Cancer J Clin*. 2004; 54(5):273-80.
4. Kingsley BS, Schmeichel KL, Rubin CH. *Environ Health Perspect.* 2007; 115:165–171.
5. National Cancer Institute. Accessed: 2/2/07 at http://cis.nci.nih.gov/fact/3_58.htm.